Replacing Cliques by Stars in Quasi-median Graphs

Katharina T. Huber, Vincent Moulton, and Charles Semple

Department of Mathematics and Statistics
University of Canterbury
Private Bag 4800
Christchurch, New Zealand

Report Number: UCDMS2002/4 March 2002

Keywords: split, (in)compatibility, strong (in)compatibility, Buneman graph, mu split graph, quasi-median graph

REPLACING CLIQUES BY STARS IN QUASI-MEDIAN GRAPHS

KATHARINA T. HUBER, VINCENT MOULTON, AND CHARLES SEMPLE

ABSTRACT. For a multi-set Σ of splits (bipartitions) of a finite set X, we introduce the multi-split graph $G(\Sigma)$. This graph is a natural extension of the Buneman graph. Indeed, it is shown that several results pertaining to the Buneman graph extend to the multi-split graph. In addition, in case Σ is derived from a set $\mathcal{R}$ of partitions of X by taking parts together with their complements, we show that the extremal instances where $\mathcal{R}$ is either strongly compatible or strongly incompatible are equivalent to $G(\Sigma)$ being either a tree or a Cartesian product of star trees, respectively.

Keywords: split, (in)compatibility, strong (in)compatibility, Buneman graph, multi-split graph, quasi-median graph

1. Introduction

A fundamental task in areas of classification is to find graphical representations of a set $\mathcal{R}$ of partitions of a finite set X. For example, in evolutionary biology X may be a set of species and the elements of $\mathcal{R}$ may be induced by given characters on X (where a pair of elements of X are in the same part of the partition associated to some character if and only if the character assigns them the same state), and the biologist often seeks to represent $\mathcal{R}$ by a tree.

In case $\mathcal{R}$ consists of bipartitions or *splits* of X that satisfy a certain pairwise property there is a very natural graphical representation which we now recall. We say that a pair $A_1|B_1$ and $A_2|B_2$ of splits of X is *compatible* if at least one of the unions $A_1 \cup A_2$, $A_1 \cup B_2$, $B_1 \cup A_2$, and $B_1 \cup B_2$ equals X (where, throughout this paper, we denote any partition $\{A_1, A_2, \ldots, A_k\}$ of a finite set by $A_1|A_2|\cdots|A_k$). Buneman [4] showed that $\mathcal{R}$ is a set of pairwise compatible splits of X if and only if $\mathcal{R}$ can be represented by a canonical tree whose leaves are labelled by the elements of X and whose edges display the elements of $\mathcal{R}$.

In practice, however, if $\mathcal{R}$ is an arbitrary set of partitions of X the most appropriate way to represent $\mathcal{R}$ is far from clear, even for "well-behaved" sets of partitions as we now illustrate. In [5], it was shown that a set $\mathcal{R}$ of partitions can be represented by a tree in case $\mathcal{R}$ is *strongly compatible* (i.e. for every distinct pair

Date: March 22, 2002.

The first and second authors were supported by the Swedish Research Council (VR) and Linnaeus Center for Bioinformatics. All authors were supported by the Swedish Foundation for International Cooperation in Research and Education (STINT) and the New Zealand Marsden Fund.

of partitions $P, Q \in \mathcal{R}$, there exists some A in P and B in Q with $A \cup B = X$). Consider the following two methods that can be used to construct such a tree.

(I) Associate the set $\Sigma_P = \{A|X - A : A \in P\}$ of splits of X to each partition P in $\mathcal{R}$. Since every pair of splits in $\Sigma_{\mathcal{R}} = \bigcup_{P \in \mathcal{R}} \Sigma_P$ is necessarily compatible, we can represent $\mathcal{R}$ by the canonical tree associated to $\Sigma_{\mathcal{R}}$ mentioned above.

(II) Associate to $\mathcal{R}$ the quasi-median graph $Q_{\mathcal{R}}$ [1]. By [1, Theorem 2] $Q_{\mathcal{R}}$ must be a block graph (i. e. every maximal 2–connected subgraph of $Q_{\mathcal{R}}$ is a clique). Hence, we can represent $\mathcal{R}$ by the tree obtained through replacing every maximal clique of size $n \geq 2$ in $Q_{\mathcal{R}}$ with a star tree of size n (i.e. a tree on n leaves with exactly one non-leaf vertex) in which the leaves of the latter are identified with the vertices of the former.

Clearly one would hope that the trees constructed by methods (I) and (II) should be the same. But, as the following example shows, this is not necessarily the case. Consider the strongly compatible set $\mathcal{R} = \{12|34, 1|2|34, 12|3|4\}$ of partitions. The tree associated to $\Sigma_{\mathcal{R}}$, the quasi-median graph $Q_{\mathcal{R}}$, and the tree obtained from $Q_{\mathcal{R}}$ by replacing every maximal clique with a star tree of the appropriate size is shown in Figure 1(a), (b), and (c), respectively.

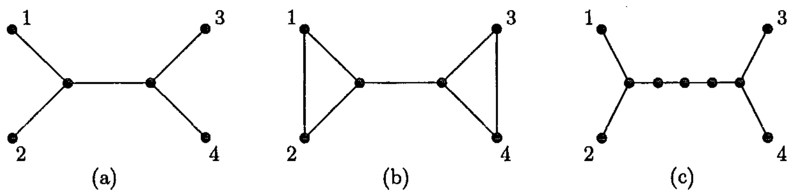


FIGURE 1

Motivated by this fact, we observed that we could shed light on the inconsistency of methods (I) and (II) by considering what we call the *multi-split graph*. As we shall see, this graph is a natural generalization of the so-called Buneman graph of a *set* of splits, a graph that for a set of pairwise compatible splits is precisely Buneman's canonical tree. In Theorem 4.4 we show that if $\mathcal{R}$ is a set of strongly compatible partitions of X, then the multi-split graph $G(\Sigma_{\mathcal{R}})$ with $\Sigma_{\mathcal{R}}$ regarded as a *multi-set* is a tree that canonically represents $\mathcal{R}$. As a consequence of this theorem and results in [1], it follows that, in case $\mathcal{R}$ is strongly compatible, $G(\Sigma_{\mathcal{R}})$ is precisely the tree obtained by method (II).

Intriguingly, this is not the only instance in which $G(\Sigma_{\mathcal{R}})$ can be obtained from $Q_{\mathcal{R}}$ by the replacement process described in (II). In the final section, we show that $G(\Sigma_{\mathcal{R}})$ is always a subgraph of a graph $St_{\mathcal{R}}$ that is isomorphic to the "Cartesian product" $\prod_{P \in \mathcal{R}} St_{|P|}$ of star trees where St_n denotes the star tree of size n. For example, suppose $\mathcal{R}$ consists of the two partitions $1|23|45$ and $125|34$ on the set $\{1, 2, 3, 4, 5\}$. Then the quasi-median graph $Q_{\mathcal{R}}$, which is isomorphic to $K_2 \times K_3$, and the multi-split graph $G(\Sigma_{\mathcal{R}})$ is shown in Figure 2(a) and (b), respectively. In this example, we clearly see that $G(\Sigma_{\mathcal{R}})$ is a subgraph of $St_2 \times St_3$.

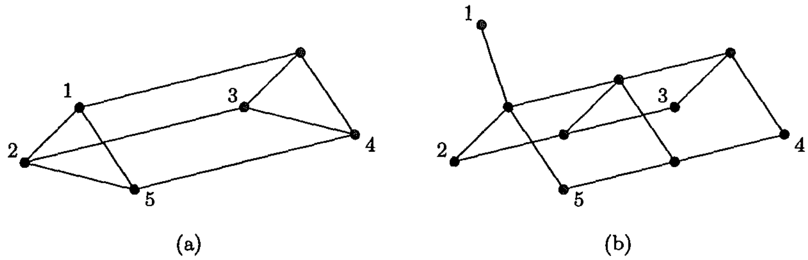


FIGURE 2

Thus, it is desirable to characterize those sets of partitions $\mathcal{R}$ for which $G(\Sigma_{\mathcal{R}})$ and $St_{\mathcal{R}}$ coincide. In Theorem 5.3 we do precisely this, proving, for a set $\mathcal{R}$ of partitions of X, that $G(\Sigma_{\mathcal{R}})$ equals $St_{\mathcal{R}}$ if and only if $\mathcal{R}$ is *strongly incompatible* (i.e. for every distinct pair of partitions $P, Q \in \mathcal{R}$, the intersection $A \cap B$ is non-empty for all $A \in P$ and $B \in Q$). Since $\mathcal{R}$ being strongly incompatible implies that $Q_{\mathcal{R}}$ is isomorphic to the Hamming graph $\prod_{P \in \mathcal{R}} K_{|P|}$ [1, Corollary 1], it follows for such $\mathcal{R}$ that $G(\Sigma_{\mathcal{R}})$ can be obtained from $Q_{\mathcal{R}}$ by simply replacing each clique $K_{|P|}$ in the expression $\prod_{P \in \mathcal{R}} K_{|P|}$ by the star tree $St_{|P|}$.

The paper is structured as follows. In Section 2, we outline some concepts of multi-sets and graphs that are needed in this paper. In Section 3, we formally define the multi-split graph and obtain various properties of it. Theorems 4.4 and 5.3 are proven in Sections 4 and 5, respectively. Furthermore, in Section 4, we describe the way in which a set of partitions of X is represented by the associated multi-split graph; this is stated as Theorem 4.2.

2. PRELIMINARIES

Throughout this paper, X denotes a finite set, and a partition of a set contains at least two parts. Both multi-sets and graphs play an important role in this paper. We briefly outline some concepts that we use about these objects.

A multi-set is like a set except that it can contain repeated elements. For example, $M_1 = \{a, a, b, b, b, c\}$ is a multi-set in which the elements a and b are repeated two and three times, respectively. Clearly, a multi-set in which each element is repeated precisely once can be regarded as a set and vice-versa. Given a multi-set M, we call the set in which repeats of elements are all removed from M the *underlying set* of M and denote it by $\underline{M}$. Thus, for the multi-set M_1, we have $\underline{M}_1 = \{a, b, c\}$. Given two multi-sets M and M', we define the difference $M - M'$ of M and M' to be the multi-set containing the elements from M each of which is repeated the number of times it occurs in M less the number of times it occurs in M' where, of course, if this difference is non-positive, the element is ignored. For example, if $M_2 = \{a, a, a, b, b\}$ then $M_1 - M_2 = \{b, c\}$. In addition, we define the union

$M \cup M'$ to be the multi-set containing the elements from M and M' each of which is repeated the number of times it occurs in M plus the number of times it occurs in M'. Hence $M_1 \cup M_2 = \{a, a, a, a, a, b, b, b, b, b, c\}$. Note that we use the same symbol for set union and multi-set union, and only explicitly state which one we are using in case it is not clear from the context. The *symmetric difference* $M \Delta M'$ of the multi-sets M and M' is the multi-set $(M - M') \cup (M' - M)$.

As usual, a *graph* is a pair $G = (V, E)$ consisting of a finite set $V = V(G)$ of vertices, together with an edge set $E = E(G)$ of 2-element subsets of V. A graph H is a *subgraph* of a graph G if both $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$ hold. If V' is a subset of $V(G)$, then the subgraph of G whose vertex set is V' and whose edge set consists of those edges in G that have both end-vertices in V' is called the subgraph of G *induced* by V'. Furthermore, the graph $G \backslash V'$ is the graph obtained from G by deleting the vertices in V' and their incident edges. If E' is a subset of $E(G)$, the graph $G \backslash E'$ is the graph obtained from G by deleting the edges in E'. In case $V = \{v\}$, we will write $G \backslash v$ rather than $G \backslash \{v\}$.

For all $i \in \{1, 2, \ldots, n\}$, let $G_i = (V_i, E_i)$ be a graph. The *(Cartesian) product* of $G_1, G_2, \ldots, G_n$, denoted $\prod_{i \in \{1,\ldots,n\}} G_i$, is the graph that has vertex set $V_1 \times \cdots \times V_n$, with an edge joining two vertices $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ precisely if, for some $i \in \{1, \ldots, n\}$, we have $\{a_i, b_i\} \in E_i$ and $a_j = b_j$ for all $j \in \{1, 2, \ldots, n\} - \{i\}$.

Lastly, let P be a partition of X, $G = (V, E)$ a graph, and $\phi : X \to V$ a map. Denoting G together with the map ϕ by $(G; \phi)$, a subset V' of V (respectively, E' of E) *displays P in G* if, for all distinct parts A and B of P, $\phi(A)$ and $\phi(B)$ are subsets of the vertex sets of distinct components of $G \backslash V'$ (respectively, $G \backslash E'$). In case $V' = \{v\}$, we say that v displays P.

3. The Multi-Split Graph

To define the multi-split graph for a multi-set Σ of splits of X or *X-splits*, we first need to describe a particular type of map associated with Σ. A *Σ-map* is a map $\psi : \Sigma \to 2^X$ such that, for all $\sigma, \sigma' \in \Sigma$,

(S1) $\psi(\sigma) \in \sigma$, and
(S2) if $\psi(\sigma) \cap \psi(\sigma')$ is empty, then $\psi(\sigma) \cup \psi(\sigma') = X$.

Given a Σ-map ψ, we denote the multi-set $\{\psi(\sigma) : \sigma \in \Sigma\}$ by $\overline{\psi}$.

The *multi-split graph on* Σ, denoted $G(\Sigma)$, is defined as follows. The vertex set $V(\Sigma)$ of $G(\Sigma)$ is the <u>set</u> $\{\overline{\psi} : \psi$ is a Σ-map$\}$. The edge set $E(\Sigma)$ of $G(\Sigma)$ consists of all 2-element subsets $\{\overline{\psi}, \overline{\omega}\}$ of $V(\Sigma)$ with

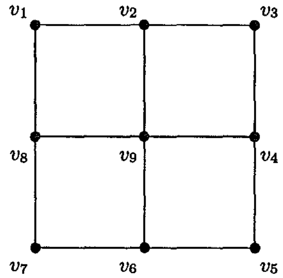
$$|\overline{\psi} \Delta \overline{\omega}| = 2.$$

Note that, if Σ has no repeated elements, that is, Σ is a set, then $G(\Sigma)$ is the familiar *Buneman graph* on Σ (see [3] for a definition of this graph).

To illustrate the multi-split graph, suppose that Σ is the multi-set

$$\{12|34, 12|34, 13|24, 13|24\}.$$

Then $G(\Sigma)$ is the graph shown in Figure 3.



$$\begin{array}{ll}
v_1 = \{\{1,2\},\{1,2\},\{1,3\},\{1,3\}\} & v_6 = \{\{3,4\},\{1,2\},\{2,4\},\{2,4\}\} \\
v_2 = \{\{1,2\},\{3,4\},\{1,3\},\{1,3\}\} & v_7 = \{\{1,2\},\{1,2\},\{2,4\},\{2,4\}\} \\
v_3 = \{\{3,4\},\{3,4\},\{1,3\},\{1,3\}\} & v_8 = \{\{1,2\},\{1,2\},\{2,4\},\{1,3\}\} \\
v_4 = \{\{3,4\},\{3,4\},\{1,3\},\{2,4\}\} & v_9 = \{\{1,2\},\{3,4\},\{2,4\},\{1,3\}\} \\
v_5 = \{\{3,4\},\{3,4\},\{2,4\},\{2,4\}\} &
\end{array}$$

FIGURE 3

Some of the notions associated with the Buneman graph on a set of splits can be easily extended to the multi-set graph on a multi-set of splits. For example, for a multi-set Σ of X–splits, let $\phi_\Sigma : X \to V(\Sigma)$ be the map defined, for all $x \in X$, by putting $\phi_\Sigma(x)$ equal to the necessarily unique vertex of $G(\Sigma)$ in which every element contains x. Furthermore, suppose that $\sigma = A|B$ is a split in Σ. Then two vertices $\overline{\psi}$ and $\overline{\omega}$ of $G(\Sigma)$ *disagree* on σ if both A and B are elements of $\overline{\psi} \,\Delta\, \overline{\omega}$, and an edge of $G(\Sigma)$ *represents* σ if its end-vertices disagree on σ. We denote the set of edges of $G(\Sigma)$ that represent σ by $E(\sigma)$.

Let Σ be a multi-set of splits of X. We next describe a recursive process that constructs a graph from the Buneman graph on $\underline{\Sigma}$. As we shall soon see, the resulting graph is the multi-split graph on Σ.

Arbitrarily order the elements of the multi-set $\Sigma - \underline{\Sigma}$ as $\sigma_1, \sigma_2, \ldots, \sigma_n$. Set G_0 to be the Buneman graph $G(\underline{\Sigma})$. For all $i \in \{1, 2, \ldots, n\}$, let G_i be the graph obtained from G_{i-1} by performing the following sequence of operations.

(I) Choose any element, B_i say, of σ_i and, for each edge e of G_{i-1} that represents σ_i and has the property that one end-vertex, $\overline{\psi}_i$ say, does not contain B_i and the other end-vertex contains exactly one B_i, subdivide e and insert the new vertex $\overline{\psi}_i \cup \{B_i\}$.

(II) For each pair of new vertices $\overline{\psi}_i \cup \{B_i\}$ and $\overline{\omega}_i \cup \{B_i\}$ such that $\overline{\psi}_i$ and $\overline{\omega}_i$ are adjacent in G_{i-1}, add an edge joining $\overline{\psi}_i \cup \{B_i\}$ and $\overline{\omega}_i \cup \{B_i\}$.

(III) Lastly, replace each non-new vertex $\overline{\psi}$, that is, each vertex in G_{i-1}, by
(a) $\overline{\psi} \cup \{A_i\}$ if $B_i \notin \overline{\psi}$ where $A_i = X - B_i$, and
(b) $\overline{\psi} \cup \{B_i\}$ otherwise.

The next proposition shows that G_n is equal to $G(\Sigma)$.

Proposition 3.1. *The graph G_n constructed above is equal to $G(\Sigma)$. Moreover, suppose that $\sigma = A|B$ is repeated exactly k times in Σ. Then $G(\Sigma)$ has the following properties:*

(i) *Each component of the subgraph of $G(\Sigma)$ that has vertex set consisting of the end-vertices of each edge in $E(\sigma)$ and edge set $E(\sigma)$ is a path $\overline{\psi}_1, \overline{\psi}_2, \ldots, \overline{\psi}_{k+1}$ consisting of k edges. Moreover, up to labelling, A appears exactly $k-(j-1)$ times in $\overline{\psi}_j$ for all $j \in \{1, 2, \ldots, k+1\}$.*
(ii) *For all $j \in \{1, 2, \ldots, k\}$, the subset of $E(\sigma)$ consisting of those edges $\{\overline{\psi}, \overline{\omega}\}$ that have the property that A appears exactly j times in $\overline{\psi}$ and $j-1$ times in $\overline{\omega}$ is a minimal set of edges of $G(\Sigma)$ displaying σ.*

Proof. The proof uses induction on n to simultaneously prove the correctness of the construction as well as (i) and (ii). The fact that the result holds for $n = 0$ follows from results in [3]. Now assume that $n \geq 1$ and that the entire proposition holds for $n-1$, in particular, G_{n-1} is equal to $G(\Sigma - \{\sigma_n\})$. It easily follows from the induction assumption that G_n satisfies (i) and (ii). Thus, to complete the proof, it suffices to show that G_n is equal to $G(\Sigma)$.

Evidently, it follows by our induction assumption and the fact that only edges that represent σ_n are subdivided in the construction process that G_n is a subgraph of $G(\Sigma)$. Thus it suffices to show that both $V(\Sigma) \subseteq V(G_n)$ and $E(\Sigma) \subseteq E(G_n)$ hold.

We first show that $V(\Sigma) \subseteq V(G_n)$ holds. Let $\overline{\omega}$ be a vertex of $G(\Sigma)$. Since $\sigma_n = A_n|B_n$ is an element of Σ, it follows that $\overline{\omega}$ can be obtained from a vertex $\overline{\psi}$ of $G(\Sigma - \{\sigma_n\}) = G_{n-1}$ by adding A_n to $\overline{\psi}$ if $A_n \in \overline{\psi}$ and adding B_n to $\overline{\psi}$ if $B_n \in \overline{\psi}$. A routine check using the induction assumption shows that all such vertices are in $V(G_n)$.

Next we show that $E(\Sigma) \subseteq E(G_n)$ holds. By the construction of the vertex set of G_n, the only possible case where there may be an edge f in $E(\Sigma)$ that is not in $E(G_n)$ is when one end-vertex of f does not, for some split $\sigma_i = A_i|B_i$ in Σ, contain B_i and the other end-vertex does not contain A_i. Since σ_i must occur at least twice in Σ, these two end-vertices disagree on at least two splits which is impossible. This completes the proof of the proposition. □

In reference to the above construction, we call, for all $i \in \{1, \ldots, n\}$, one iteration of this construction a *parallel subdivision of G_{i-1} on σ_i*. The next corollary is an immediate consequence of Lemma 3.1.

Corollary 3.2. *Let Σ be a multi-set of splits of X. Then $G(\Sigma)$ can be obtained from the Buneman graph on $\underline{\Sigma}$ by a sequence of parallel subdivisions.*

In light of Corollary 3.2, many basic properties of the Buneman graph on $\underline{\Sigma}$ can be easily seen to extend to $G(\Sigma)$. The next two results illustrate this.

A connected graph G is a *median* graph if, for every three vertices u_1, u_2, and u_3 of G, there is exactly one vertex v of G that simultaneously lies on shortest paths joining u_1 and u_2, u_1 and u_3, and u_2 and u_3. Since the Buneman graph on a set of splits is connected [2], it follows by Corollary 3.2 that the multi-split graph on a multi-set of splits is also connected. Following the proof [2, Theorem 1] that the Buneman graph is a median graph, one immediately obtains the next proposition.

Proposition 3.3. *Let Σ be a multi-set of splits of X. Then $G(\Sigma)$ is a median graph.*

Recall that two X–splits $A|B$ and $A'|B'$ are *compatible* if at least one of the four intersections $A \cap A'$, $A \cap B'$, $B \cap A'$, and $B \cap B'$ is empty. A multi-set Σ of X–splits is called *pairwise compatible* if every pair of splits in Σ is compatible. Evidently, Σ is pairwise compatible if and only if $\underline{\Sigma}$ is pairwise compatible.

A *semi-labelled tree on X* is an ordered pair $(T; \phi)$, where T is a tree with vertex set V and $\phi : X \to V$ is a map with the property that all leaves in T are contained in $\phi(X)$. For those readers familiar with X–trees, an *X-tree* has the additional property that all vertices of degree two are also contained in $\phi(X)$.

Proposition 3.4. *Let Σ be a multi-set of splits of X. Then the following statements hold:*

(i) *The multi-split graph $G(\Sigma)$ is a tree if and only if Σ is pairwise compatible.*
(ii) *If Σ is pairwise compatible, then the multi-set of X-splits displayed by the edges of $(G(\Sigma); \phi_\Sigma)$ equals Σ. Moreover, $(G(\Sigma); \phi_\Sigma)$ is the only semi-labelled tree on X with this property.*

Proof. We first prove (i). Clearly, $G(\underline{\Sigma})$ is a tree if and only if $G(\Sigma)$ is a tree since the construction of $G(\Sigma)$ from $G(\underline{\Sigma})$ by a sequence of parallel subdivisions described above introduces no cycles. As (i) holds in the case Σ is a set of X–splits (see [3]), it follows that (i) holds if Σ is a multi-set of X–splits.

To see that (ii) holds, first note that it holds in case Σ is a set (see [3]). Combining this fact with part (i) and Proposition 3.1, we deduce that part (ii) holds for a multi-set of splits of X. □

4. $(X, \mathcal{R})$–Trees

Let $\mathcal{R}$ be a set of partitions of X and let P an element in $\mathcal{R}$. Recall that the multi-set $\{A|X - A : A \in P\}$ is denoted by Σ_P, and the multi-set $\bigcup_{P \in \mathcal{R}} \Sigma_P$ is denoted by $\Sigma_{\mathcal{R}}$. We first show that, for all elements P in $\mathcal{R}$, there is a canonical set of vertices of $G(\Sigma_{\mathcal{R}})$ that displays P. To establish this result, we make use of the following lemma whose straightforward proof is omitted.

Lemma 4.1. *Let $\mathcal{R}$ be a set of partitions of X, and suppose that ψ is a $\Sigma_{\mathcal{R}}$-map. Then ψ has the property that, for all non-bipartitions $P \in \mathcal{R}$,*

$$|P \cap \psi(\Sigma_P)| \leq 1.$$

Theorem 4.2. *Let $\mathcal{R}$ be a set of partitions of X, and let P be an element of $\mathcal{R}$. Then the subset V_P of $V(\Sigma_{\mathcal{R}})$ given by*

$$V_P = \{\overline{\psi} : \overline{\psi} \in V(\Sigma_{\mathcal{R}}) \text{ and } X - A \in \overline{\psi} \text{ for all } A \in P\}$$

displays P in $(G(\Sigma_{\mathcal{R}}); \phi_{\Sigma_{\mathcal{R}}})$.

Proof. Suppose $P = A_1|A_2|\ldots|A_k$, where $k \geq 2$. Consider the pair $(G(\Sigma_{\mathcal{R}}); \phi_{\Sigma_{\mathcal{R}}})$. For all $i \in \{1, 2, \ldots, k\}$, let $E(A_i)$ denote the subset of $E(\Sigma_{\mathcal{R}})$ consisting of those edges that represent $A_i|X - A_i$ and have the property that one end-vertex does not contain $X - A_i$. By Proposition 3.1, for all i, $E(A_i)$ is a minimal set of edges of $G(\Sigma_{\mathcal{R}})$ that displays $A_i|X - A_i$. For all i, let G_i denote the component of $G(\Sigma_{\mathcal{R}})\backslash E(A_i)$ that displays A_i. Then, for all i, no vertex of G_i contains $X - A_i$, but every vertex in $V(\Sigma_{\mathcal{R}}) - V(G_i)$ contains $X - A_i$, and so $\phi_{\Sigma_{\mathcal{R}}}(A_i) \subseteq V(G_i)$. It is easily seen that, for $|P| = 2$, the intersection $V(G_1) \cap V(G_2)$ is empty. Furthermore, it follows, using Lemma 4.1 for the case $|P| \geq 3$, that $V(G_i) \cap V(G_j)$ is empty for all distinct $i, j \in \{1, 2, \ldots, k\}$. Thus $G(\Sigma_{\mathcal{R}})$ is of the form shown in Figure 4. It is now easily seen that V_P is the set

$$V(\Sigma_{\mathcal{R}}) - \bigcup_{i \in \{1, \ldots, k\}} V(G_i),$$

and so V_P does indeed display P. □

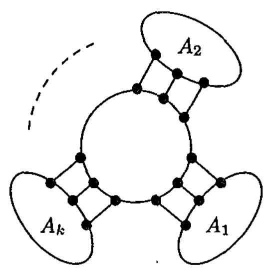


FIGURE 4

Recall that two partitions P and Q of X are called *strongly compatible* if either $P = Q$ or there is an element A in P and an element B in Q with $A \cup B = X$. A set $\mathcal{R}$ of partitions of X is *strongly compatible* if every pair of partitions in $\mathcal{R}$ is strongly compatible. The following lemma relates pairwise compatibility to strong compatibility. Its straightforward proof is omitted.

Lemma 4.3. *Let $\mathcal{R}$ be a set of partitions of X. Then the following statements are equivalent.*

(i) *$\Sigma_{\mathcal{R}}$ is pairwise compatible.*

(ii) *For all pairs of partitions $P, Q \in \mathcal{R}$, either P and Q are strongly compatible, or every element A of the set $P \cup Q$ is the disjoint union of elements in the multi-set $(P \cup Q) - A$.*

For a set $\mathcal{R}$ of partitions of X, an *$(X, \mathcal{R})$–tree* $(T; \phi; \kappa)$ is a semi-labelled tree $(T; \phi)$, where $T = (V, E)$, together with an additional labelling map $\kappa : \mathcal{R} \to V$ with the property that, for all $P \in \mathcal{R}$, the vertex $\kappa(P)$ displays P.

The next theorem can be viewed as an extension of Proposition 3.4 from bipartitions to partitions.

Theorem 4.4. *Let $\mathcal{R}$ be a set of partitions of X. Then there exists an $(X, \mathcal{R})$–tree if and only if $\mathcal{R}$ is strongly compatible. Moreover, if such a tree exists, then, up to isomorphism and choice of vertex that displays a bipartition in $\mathcal{R}$, there is a unique $(X, \mathcal{R})$–tree for which the multi-set of splits displayed by its edges is equal to $\Sigma_{\mathcal{R}}$, and this tree is precisely $G(\Sigma_{\mathcal{R}})$.*

Proof. Suppose that there exists an $(X, \mathcal{R})$–tree $(T; \phi; \kappa)$. Let P_1 and P_2 be elements of $\mathcal{R}$, and let v_1 and v_2 be the vertices of this tree displaying P_1 and P_2, respectively. Let A_1 be the part of P_1 displayed by the component of $T \backslash v_1$ that contains v_2, and let A_2 be the part of P_2 displayed by the component of $T \backslash v_2$ that contains v_1. It is easily seen that $A_1 \cup A_2 = X$ holds. It follows that $\mathcal{R}$ is indeed strongly compatible.

Now suppose $\mathcal{R}$ is strongly compatible and consider the pair $(G(\underline{\Sigma}_{\mathcal{R}}); \phi_{\underline{\Sigma}_{\mathcal{R}}})$ where we write $\underline{\Sigma}_{\mathcal{R}}$ instead of $\underline{\Sigma_{\mathcal{R}}}$. By Lemma 4.3, $\Sigma_{\mathcal{R}}$ is a pairwise compatible multi-set of X–splits.

4.4.1. *If Q is an element of $\mathcal{R}$ with $|Q| \geq 3$, then there is a unique vertex of $G(\underline{\Sigma}_{\mathcal{R}})$ that displays Q.*

Proof. Since $\underline{\Sigma}_{\mathcal{R}}$ is pairwise compatible, $(G(\underline{\Sigma}_{\mathcal{R}}); \phi_{\underline{\Sigma}_{\mathcal{R}}})$ is an X–tree (see [3]). In particular, every degree-two vertex u of $G(\underline{\Sigma}_{\mathcal{R}})$ is contained in $\phi_{\underline{\Sigma}_{\mathcal{R}}}(X)$. Suppose $Q = \{A_1, A_2, \ldots, A_k\}$, where $k \geq 3$. Then, for all $i \in \{1, 2, \ldots, k\}$, it follows by Proposition 3.4 that there is a unique edge e_i of $G(\underline{\Sigma}_{\mathcal{R}})$ that displays $A_i | X - A_i$. Since $A_1, A_2, \ldots, A_k$ partitions X, it is now easily seen that either $e_1, e_2, \ldots, e_k$ are incident with a common vertex v, in which case, v displays Q and no other vertex has this property, or $G(\underline{\Sigma}_{\mathcal{R}})$ is of the form shown in Figure 5 where $j \geq 2$ and $k - r \geq 2$ both hold. In the latter case, it again follows by Proposition 3.4 that $A_1 \cup \cdots \cup A_j | X - (A_1 \cup \cdots \cup A_j)$ is an X–split contained in $\underline{\Sigma}_{\mathcal{R}}$, and so one part of this split is a part of a partition P in $\mathcal{R}$. Without loss of generality, we may assume that this part is $Y = X - (A_1 \cup \cdots \cup A_j)$. Since $j \geq 2$, there is no element A_i of Q such that $Y \cup A_i = X$. Furthermore, as $k - r \geq 2$, there is no element B of $P - Y$ such that $B \cup A_i = X$. This implies that P and Q are not strongly compatible; a contradiction. Thus there is a unique vertex of $G(\underline{\Sigma}_{\mathcal{R}})$ that displays Q. $\square$

We now complete the proof of the theorem. Since $G(\underline{\Sigma}_{\mathcal{R}})$ is a tree, it follows by Corollary 3.2 that the graph $G(\Sigma_{\mathcal{R}})$ can be obtained from $G(\underline{\Sigma}_{\mathcal{R}})$ by simply subdividing edges and relabelling vertices and $\phi_{\Sigma_{\mathcal{R}}}$ is induced by $\phi_{\underline{\Sigma}_{\mathcal{R}}}$. Let P be

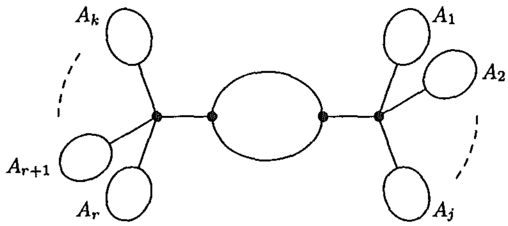


FIGURE 5

an element of $\mathcal{R}$. If $|P| \geq 3$, then (4.4.1) implies that the vertex of $G(\underline{\Sigma}_{\mathcal{R}})$ that displays P also displays P in $G(\Sigma_{\mathcal{R}})$. Moreover, this is the only such vertex that displays P.

Now assume that $|P| = 2$. Then P is a bipartition and appears at least twice in $\Sigma_{\mathcal{R}}$. Therefore, by Proposition 3.1, there are two adjacent edges of $G(\Sigma_{\mathcal{R}})$ that display P and it follows that the vertex incident with both of these edges displays P. Hence there exists an $(X, \mathcal{R})$–tree. The fact that this tree has the desired uniqueness property follows by Proposition 3.4(ii). □

Remark. In [5, Theorem 5.6], it is shown that, for a finite set X and a family χ of characters on X, there exists an "(X, χ)–tree", a semi-labelled tree analogous to an $(X, \mathcal{R})$–tree, if and only if every pair of characters in χ are strongly compatible. Since, as described in the introduction, any character on X can be naturally associated to a partition of X and vice-versa it is straight-forward to see that Theorem 3.4 provides an independent proof of this result. In addition, in case χ is strongly compatible, Theorem 4.4 provides an explicit description of the unique (X, χ)–tree associated to χ that was alluded to in [5, Remark 5.7].

5. Strong Incompatibility

For a multi-set Σ of splits of X, a *weak Σ–map* is a map from Σ to 2^X that satisfies property (S1) in the definition of a Σ–map. It is well-known that the Buneman graph on a set Σ consisting of k splits of X is a vertex induced subgraph of the *k–cube* that is, the graph with vertex set $\{\overline{\psi} : \psi \text{ is a weak } \Sigma\text{–map}\}$ and an edge joining two vertices $\overline{\psi}$ and $\overline{\omega}$ precisely if $|\overline{\psi} \,\Delta\, \overline{\omega}| = 2$. We begin this section by describing an analogous result for the multi-split graph that is derived from a set of partitions.

Let $\mathcal{R}$ be a set of partitions of X. For a weak $\Sigma_{\mathcal{R}}$–map ψ, let (S3) denote the following property (see Lemma 4.1).

(S3) For all non-bipartitions $P \in \mathcal{R}$, $|P \cap \psi(\Sigma_P)| \leq 1$.

Now let $St_{\mathcal{R}}$ denote the graph that is defined as follows. The vertex set of $St_{\mathcal{R}}$ is the set

$$\{\overline{\psi} : \psi \text{ is a weak } \Sigma_{\mathcal{R}}\text{-map satisfying (S3)}\}$$

and the edge set of $St_{\mathcal{R}}$ consists of all pairs $\{\overline{\psi}, \overline{\omega}\}$ such that $|\overline{\psi} \Delta \overline{\omega}| = 2$. The reason for the notation "$St_{\mathcal{R}}$" is that, as we shall now show, $St_{\mathcal{R}}$ is isomorphic to a product of star trees. For a partition P, let St_P denote the star tree whose set of pendant vertices equals P.

Proposition 5.1. *Let $\mathcal{R}$ be a set of partitions of X. Then $St_{\mathcal{R}}$ is isomorphic to $\prod_{P \in \mathcal{R}} St_P$.*

Proof. Let P be an element of $\mathcal{R}$. It is straightforward to check that the map Φ from the vertex set of St_P into the vertex set of $G(\Sigma_P)$ defined, for all $A \in V(St_P)$, by

$$\Phi(A) = \begin{cases} \{A\} \cup \bigcup_{B \in P - A} \{X - B\}, & \text{if } A \text{ is a pendant vertex;} \\ \bigcup_{B \in P} \{X - B\}, & \text{otherwise,} \end{cases}$$

induces an isomorphism between St_P and $G(\Sigma_P)$. Consequently, $\prod_{P \in \mathcal{R}} St_P$ is isomorphic to $\prod_{P \in \mathcal{R}} G(\Sigma_P)$. Denote the latter Cartesian product by $C_{\Sigma_{\mathcal{R}}}$. To see that $C_{\Sigma_{\mathcal{R}}}$ is isomorphic to $St_{\mathcal{R}}$, view each $|\mathcal{R}|$-tupled vertex of $C_{\Sigma_{\mathcal{R}}}$ as the multi-set that is the union of the components of the $|\mathcal{R}|$-tuple. Then the vertex set of $C_{\Sigma_{\mathcal{R}}}$ is equal to the vertex set of $St_{\mathcal{R}}$. Furthermore, under this viewpoint, e is an edge of $C_{\Sigma_{\mathcal{R}}}$ precisely if the symmetric difference of the end-vertices of e is equal to 2. Hence $C_{\Sigma_{\mathcal{R}}}$, and in particular $\prod_{P \in \mathcal{R}} St_P$, is isomorphic to $St_{\mathcal{R}}$. □

By Lemma 4.1, $V(\Sigma_{\mathcal{R}})$ is a subset of $V(St_{\mathcal{R}})$. The next proposition is an immediate consequence of this fact, and the definitions of the edge sets of $G(\Sigma_{\mathcal{R}})$ and $St_{\mathcal{R}}$.

Proposition 5.2. *Let $\mathcal{R}$ be a set of partitions of X. Then $G(\Sigma_{\mathcal{R}})$ is a vertex induced subgraph of $St_{\mathcal{R}}$.*

Note that the inclusion in Proposition 5.2 can be proper. For example, suppose that $\mathcal{R}$ consists of the partitions $12|34|56$ and $123|456$ of the set $\{1, \ldots, 6\}$. Then no vertex of $G(\Sigma_{\mathcal{R}})$ contains both $\{1, 2\}$ and $\{4, 5, 6\}$, but there is such a vertex in $St_{\mathcal{R}}$.

The main purpose of the rest of this section is to characterize sets $\mathcal{R}$ of partitions for which $G(\Sigma_{\mathcal{R}})$ equals $St_{\mathcal{R}}$.

Recall that two partitions P and Q of X are *strongly incompatible* if $A \cap B$ is non-empty for all $A \in P$ and $B \in Q$. Observe that, if P and Q are strongly incompatible, then P and Q must be distinct. A set $\mathcal{R}$ of partitions is *strongly incompatible* if every pair of partitions in $\mathcal{R}$ is strongly incompatible.

To illustrate, the partitions $135|246$ and $12|34|56$ of the set $\{1, 2, 3, 4, 5, 6\}$ are strongly incompatible. Furthermore, we note that it is straightforward to see that a pair of partitions cannot be both strongly incompatible and strongly compatible and, of course, that a pair of partitions can be neither strongly incompatible nor strongly compatible.

Theorem 5.3. *Let $\mathcal{R}$ be a set of partitions of X. Then $G(\Sigma_{\mathcal{R}})$ is equal to $St_{\mathcal{R}}$ if and only if $\mathcal{R}$ is strongly incompatible.*

The proof of Theorem 5.3 uses the following two lemmas whose proofs are routine and omitted.

Lemma 5.4. *Let $\mathcal{R} = \{P_1, P_2, \dots, P_n\}$ be a set of partitions of X. Then*

$$|V(St_{\mathcal{R}})| = (|P_1| + 1) \times (|P_2| + 1) \times \cdots \times (|P_n| + 1).$$

Lemma 5.5. *Suppose $\mathcal{R}$ is a set of strongly incompatible partitions of X. Let $\sigma_1, \sigma_2 \in \Sigma_{\mathcal{R}}$, and let $A_1 \in \sigma_1$ and $A_2 \in \sigma_2$. Then*

(i) *$A_1 \cap A_2 \neq \emptyset$ if and only if A_1 and A_2 are not parts of the same partition contained in $\mathcal{R}$.*
(ii) *$A_1 \cup A_2 = X$ if and only if $\{A_1, A_2\}$ is a partition in $\mathcal{R}$.*

Proof of Theorem 5.3. Let $\mathcal{R} = \{P_1, P_2, \dots, P_n\}$, and suppose that $G(\Sigma_{\mathcal{R}})$ is equal to $St_{\mathcal{R}}$. Then, by Lemma 5.4,

$$|V(\Sigma_{\mathcal{R}})| = (|P_1| + 1) \times (|P_2| + 1) \times \cdots \times (|P_n| + 1). \tag{1}$$

Let P and Q be distinct partitions in $\mathcal{R}$, and let $A \in P$ and $B \in Q$. To complete the "only if" part of Theorem 5.3, it suffices to show that $A \cap B$ is non-empty. By Proposition 5.2 and Equation (1), there must be a vertex $\overline{\psi}$ in $G(\Sigma_{\mathcal{R}})$ that contains both A and B. Assume that $A \cap B$ is empty. Then, by definition of the vertices of $G(\Sigma_{\mathcal{R}})$, X is the disjoint union of A and B. Therefore, if $A' \in P - A$ and $B' \in Q - B$, then $A' \subseteq B$ and $B' \subseteq A$. But, again by Proposition 5.2 and Equation (1), there is a vertex of $G(\Sigma_{\mathcal{R}})$ that contains both A' and B', and so $A' = B$ and $B' = A$. Thus $P = \{A, B\} = Q$, contradicting the assumption that P and Q are distinct, and so $A \cap B$ is indeed non-empty as required.

Now suppose that $\mathcal{R}$ is strongly incompatible. To show that $G(\Sigma_{\mathcal{R}})$ is equal to $St_{\mathcal{R}}$, it suffices, by Proposition 5.2, to show that every vertex in $St_{\mathcal{R}}$ is a vertex in $\Sigma_{\mathcal{R}}$. Combining Lemma 5.5 with the definition of the vertex set of $St_{\mathcal{R}}$, it is easily seen that this is indeed the case. This completes the proof of the theorem. □

Acknowledgement. KTH and VM would like to thank Mike Steel for his hospitality and for discussions which instigated the investigations presented here.

References

[1] H.-J. Bandelt, K. T. Huber, V. Moulton, Quasi-median graphs from sets of partitions, Discrete Applied Mathematics, to appear.

[2] J.-P. Barthélemy, From copair hypergraphs to median graphs with latent vertices, Discrete Math. 76 (1989) 9-28.

[3] J.-P. Barthélemy and A. Guénoche, Trees and proximity representations (Wiley, 1991).

[4] P. Buneman, The recovery of trees from measures of dissimilarity, In: F. R. Hodson, D. G. Kendall, and P. Tautu, ed., Mathematics in the Archaeological and Historical Sciences (Edinburgh University Press, Edinburg, 1971) 387–395.

[5] A. Dress, V. Moulton, M. Steel, Trees, taxonomy and strongly compatible characters, Adv. in Math. 19 (1997) 1-30.

Department of Biometry and Informatics, Swedish University of Agricultural Sciences, Box 7013, 750 07 Uppsala, Sweden.

E-mail address: huber@ekocent.ekon.slu.se

Linnaeus Centre for Bioinformatics, BMC, Uppsala University, Box 598, 751 24 Uppsala, Sweden.

E-mail address: vincent.moulton@lcb.uu.se

Biomathematics Research Centre, Department of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand

E-mail address: c.semple@math.canterbury.ac.nz